PB83-131235

Inference about the ED50 Using the Trimmed Spearman-Karber Procedure A Monte Carlo Investigation

Montana State Univ.
Bozeman

Prepared for

Environmental Research Lab.-Duluth, MN

1980

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before complet

1. REPORT NO. EPA-600/J-80-397	2. Journal Article
3. PB83-131235	
4. TITLE AND SUBTITLE: Inference about the ED50 using the Trimmed Spearman-Karber Procedure - A Monte Carlo Investigation	5. REPORT DATE: 1980
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S): Hamilton, M. A.	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Mathematical Sciences, Montana State University, Bozeman, Montana 59717	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO. R803950
12. SPONSORING AGENCY NAME AND ADDRESS: U.S. Environmental Protection Agency, Environmental Research Laboratory-Duluth, 6201 Congdon Boulevard, Duluth, Minnesota 55804	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE EPA/600/03

15. SUPPLEMENTARY NOTES: Commun. Statist.-Simula. Computa. B9(3): 235-254 (1980)

16. ABSTRACT

A Monte Carlo study was conducted to investigate the estimated standard errors of the estimate and the 95 percent confidence interval estimates associated with the trimmed Spearman-Karber (SK) estimators of the ED50 and the logistic model maximum likelihood estimator (MLE). The simulated binary response bioassay experiments had widely spaced doses with 5, 10, or 20 subjects per dose. For data following a logistic tolerance distribution, the trimmed SK confidence intervals were nearly as accurate as the logistic MLE intervals. For heavy-tailed tolerance distributions, the trimmed SK confidence intervals were more accurate than those based on the logistic MLE.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

18. DISTRIBUTION STATEMENT: RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)*: UNCLASSIFIED	21. NO. OF PAGES: 21
	20. SECURITY CLASS *(This page)*: UNCLASSIFIED	22. PRICE

EPA-600/J-80-397
JOURNAL ARTICLE

COMMUN. STATIST.-SIMULA. COMPUTA., B9(3), 235-254 (1980)

INFERENCE ABOUT THE ED50 USING THE TRIMMED SPEARMAN-KÄRBER PROCEDURE - A MONTE CARLO INVESTIGATION

Martin Alva Hamilton

Department of Mathematical Sciences
Montana State University
Bozeman, MT 59717

Key Words and Phrases: binary response bioassay; maximum likelihood; logistic; standard error; confidence intervals.

ABSTRACT

A Monte Carlo study was conducted to investigate the estimated standard errors of the estimate and the 95 percent confidence interval estimates associated with the trimmed Spearman-Kärber (SK) estimators of the ED50 and the logistic model maximum likelihood estimator (MLE). The simulated binary response bioassay experiments had widely spaced doses with 5, 10, or 20 subjects per dose. For data following a logistic tolerance distribution, the trimmed SK confidence intervals were nearly as accurate as the logistic MLE intervals. For heavy-tailed tolerance distributions, the trimmed SK confidence intervals were more accurate than those based on the logistic MLE.

1980 by Marcel Dekker, Inc.

1. INTRODUCTION

Consider the problem of estimating the ED50 from binary response bioassay data. In a previous paper (Hamilton (1979)), I compared a number of estimators of the ED50, including the logistic model maximum likelihood and the trimmed Spearman-Kärber estimators. The criteria of comparison were (1) accuracy and precision for logistic or contaminated logistic tolerance distributions and widely spaced dose levels; (2) sensitivity to an anomalous response; and (3) calculability. The trimmed Spearman-Kärber estimator was recommended.

Miller and Halpern (1979) have adapted L-estimators and M-estimators to the problem of estimating the ED50 from binary response bioassay data and they have investigated the large sample properties of these estimators. For each of nine tolerance distributions, Miller and Halpern calculate the asymptotic variances of the trimmed Spearman-Kärber and Tukey biweight estimators as well as the asymptotic variance of the optimum L-estimators. (The trimmed Spearman-Kärber estimator is an L-estimator, the Tukey biweight is an M-estimator, and the Spearman-Kärber estimator is both an L-estimator and an M-estimator.) They compare the asymptotic variances to the small sample Monte Carlo results of Hamilton (1979). This large sample work indicates that the trimmed Spearman-Kärber performs relatively well, especially if the tolerance distribution is contaminated or heavy-tailed.

The analysis of a binary response bioassay should usually include a confidence interval estimate and an estimate of the standard error of the estimate in addition to a point estimate of the ED50. In fact, bioassay laboratories that are conducting a large number of routine bioassays to monitor potentially toxic effluents are required by the Environmental Protection Agency to produce a confidence interval as well as a point estimate of the

ED50 for each bioassay experiment (Committee on Methods for Toxicity Tests with Aquatic Organisms (1975)). It is important that statistically valid procedures are used and that the validity is established in a realistic small sample setting.

The purpose of this paper is to describe a Monte Carlo investigation of the bias and variability of the standard error of the estimate and the level and accuracy of the associated 95 percent confidence interval estimate for the ED50. The primary goal is to investigate the trimmed Spearman-Karber procedure. The maximum likelihood procedure is included for comparison.

Section two of this paper is a description of the variance, standard error, and confidence interval estimates. Section three presents the results of the Monte Carlo study, section four discusses the design of the simulation study, and section five is the conclusion.

2. VARIANCE, STANDARD ERROR & CONFIDENCE INTERVAL ESTIMATORS

2.1 Preliminaries

The notation and estimators of Hamilton (1979) will be used and that article should be consulted for specific formulae and discussion. The binary response bioassay experiment yields the following information:

$x_1, x_2, \ldots, x_k$, which are the k doses of the drug used and which are arranged in increasing order so that $x_1 < x_2 < \ldots < x_k$;

$n_1, n_2, \ldots, n_k$, which are the numbers of subjects given doses $x_1, \ldots, x_k$, respectively;

$r_1, r_2, \ldots, r_k$, which are the numbers of subjects that respond to doses $x_1, \ldots, x_k$, respectively.

The observed proportion of response due to dose x_i is denoted by $s_i = r_i/n_i$, $i = 1, \ldots, k$. Denote the true dose-response proportion by $P(x)$ and let μ be the dose such that $P(\mu) = 0.50$. Then μ is the median effective dose, commonly called the ED50, and $P(x)$ is the tolerance distribution function for the underlying population of subjects. Let $P_1, \ldots, P_k$ denote the distribution-free maximum likelihood estimators of $P(x_1), \ldots, P(x_k)$; viz.,

$$P_i = \max_{1 \le u \le i} \min_{i \le v \le k} \{ \sum_{j=u}^{v} r_j / \sum_{j=u}^{v} n_j \}.$$

2.2 Description of Variance and Standard Error Estimators

In the sections which follow, formulas for the estimated variance of the estimate, $\hat{Var}(\hat{\mu})$, are given. The associated estimate of the standard error is taken to be the square root of the estimated variance; $\hat{SE}(\hat{\mu}) \equiv (\hat{Var}(\hat{\mu}))^{1/2}$. All of the 95 percent confidence interval estimators discussed in this paper have endpoints calculated simply by $\hat{\mu} \pm 2 \cdot \hat{SE}(\hat{\mu})$.

2.2.1 Maximum Likelihood Procedure. The maximum likelihood estimator considered in this paper is based on a logistic model for $P(x)$; viz., $P(x) = [1 + \exp\{-(\alpha + \beta x)\}]^{-1}$. For this logistic model, $\mu = -\alpha/\beta$. Normal equations and interative calculation procedures for finding the maximum likelihood estimate (m.l.e.) of μ are presented in Hamilton (1979). Let $\hat{P}_i = [1 + \exp\{-(\hat{\alpha} + \hat{\beta} x_i)\}]^{-1}$, where $\hat{\alpha}$ and $\hat{\beta}$ are the m.l.e. of α and β. The usual estimated covariance matrix for $(\hat{\alpha}\ \hat{\beta})'$ is (Cox 1970, pp. 87-88; Ashton 1972, section 4.4.).

$$\begin{pmatrix} \hat{Var}(\hat{\alpha}) & \hat{Cov}(\hat{\alpha},\hat{\beta}) \\ \hat{Cov}(\hat{\alpha},\hat{\beta}) & \hat{Var}(\hat{\beta}) \end{pmatrix} = \begin{pmatrix} \Sigma\hat{P}_i(1-\hat{P}_i)n_i & \Sigma x_i\hat{P}_i(1-\hat{P}_i)n_i \\ \Sigma x_i\hat{P}_i(1-\hat{P}_i)n_i & \Sigma x_i^2\hat{P}_i(1-\hat{P}_i)n_i \end{pmatrix}^{-1} \quad (2.1)$$

The estimated variance of $\hat{\mu} = -\hat{\alpha}/\hat{\beta}$ is

$$\hat{Var}(\hat{\mu}) = [\hat{Var}(\hat{\alpha}) + 2\hat{\mu}\hat{Cov}(\hat{\alpha},\hat{\beta}) + \hat{\mu}^2\hat{Var}(\hat{\beta})]/\hat{\beta}^2. \quad (2.2)$$

This estimate of the variance can be derived by using the δ-method for finding the asymptotic variance of a ratio of random variables and then by substituting estimates for the unknown parameters that occur in the asymptotic variance expression. The abbreviation MLE is used below to identify standard error and interval estimates based on (2.1) and (2.2).

2.2.2 Trimmed Spearman-Kärber Procedure. The trimmed Spearman-Kärber estimator is described in Hamilton (1979) and in Hamilton, Russo, and Thurston (1977). Let A be a chosen constant, $0.0 \leq A \leq 0.5$. Let

$L(A) = \max\{i : p_i \leq A, \ i = 1,\ldots,k\}$ and

$U(A) = \min\{i : p_i \geq 1 - A, \ i = 1,\ldots,k\}$. Let

$\hat{P}^{-1}(A) = X_{L(A)} + (X_{L(A)+1} - X_{L(A)})(A - p_{L(A)})/(p_{L(A)+1} - p_{L(A)})$

and

$\hat{P}^{-1}(1 - A) = X_{U(A)-1} + (X_{U(A)} - X_{U(A)-1})\ (1 - A - p_{U(A)-1})/(p_{U(A)} - p_{U(A)-1})$. Then the 100A percent trimmed Spearman-Kärber estimator of μ is

$$\hat{\mu}(A) = [(\hat{P}^{-1}(A) + x_{L(A)+1})\ (p_{L(A)+1} - A)/2$$
$$+ \sum_{L(A)+1}^{U(A)-2} (x_i + x_{i+1})\ (p_{i+1} - p_i)/2$$
$$+ (\hat{P}^{-1}(1 - A) + x_{U(A)-1})\ (1 - A - p_{U(A)-1})/2]/(1 - 2A),$$
$$A \neq 0.5;$$
$$= \hat{P}^{-1}(0.5), \qquad A = 0.5 \tag{2.3}$$

Suppressing the argument A of L(A) and U(A), define

$$V_1 = [(x_{L+1} - x_L)\ (p_{L+1} - A)^2/(p_{L+1} - p_L)^2]^2 p_L(1 - p_L)/n_L ;$$

$$V_2 = [(x_L + x_{L+2}) + (x_{L+1} - x_L)\ (A - p_L)^2/(p_{L+1} - p_L)^2]^2 p_{L+1}$$
$$\cdot (1 - p_{L+1})/n_{L+1};$$

$$V_3 = \sum_{L+2}^{U-2} (x_{i-1} - x_{i+1})^2\ p_i(1 - p_i)/n_i ;$$

$$V_4 = [(x_{U-2} - x_U) + (x_U - x_{U-1})\ (p_U - 1 + A)^2/$$
$$(p_U - p_{U-1})^2]^2 p_{U-1}(1 - p_{U-1})/n_{U-1} ;$$

$$V_5 = [(x_U - x_{U-1})\ (1 - A - p_{U-1})^2/(p_U - p_{U-1})^2]^2 p_U(1 - p_U)/n_U ;$$

$$V_6 = [(x_U - x_{L+1})\ (1 - A - p_U)^2/(p_U - p_{L+1})^2] -$$
$$[(x_{L+1} - x_L)\ (A - p_L)^2/(p_{L+1} - p_L)^2] +$$
$$(x_L - x_U))^2 p_{L+1}(1 - p_{L+1})/n_{L+1}. \tag{2.4}$$

An estimate of the large scale sample variance of $\hat{\mu}(A)$ is

$$\begin{aligned}
\mathrm{Var}(\hat{p}(A)) &= (V_1 + V_2 + V_3 + V_4 + V_5)/[2(1 - 2A)]^2, \\
&\quad \text{if } A<.5,\ U-L\geq 4; \\
&= (V_1 + V_2 + V_4 + V_5)/[2(1 - 2A)]^2, \\
&\quad \text{if } A<.5,\ U-L=3; \\
&= (V_1 + V_3 + V_6)/[2(1 - 2A)]^2, \\
&\quad \text{if } A<.5,\ U-L=2; \\
&= (x_U - x_L)^2 \{[(0.5 - p_U)^2/(p_U - p_L)^4]\, p_L(1 - p_L)/n_L \\
&\quad + [(0.5 - p_L)^2/(p_U - p_L)^4]\, p_U(1 - p_U)/n_U\} \\
&\quad \text{if } A=0.5 \text{ or } U-L=1. \qquad (2.5)
\end{aligned}$$

This variance estimate was derived in three steps. First, it was assumed that $P(x_L) \leq A$ and $P(x_U) \geq 1 - A$. Second, the δ-method was used to calculate the asymptotic (as $k \to \infty$ and $n_i' s \to \infty$) variance of $\hat{p}(A)$. Finally, the p_i's were substituted for the $P(x_i)$'s in the asymptotic variance formula. This is similar to the derivation employed by Bennett (1952), who proposed a variance estimate for the moving average estimate of the ED50.

If $A = 0$, $p_1 = 0$, and $p_k = 1.0$, (2.3) defines the conventional SK estimator and (2.5) defines the usual variance estimate for the conventional SK estimator as presented in Miller (1973). Standard error and confidence interval estimates based on (2.4) and (2.5) will be denoted by SK when $A = 0.0$ and by 5%SK, 10%SK, and 20%SK for $A = 0.05$, 0.10, and 0.20, respectively. No other trimming coefficients were included in the Monte Carlo study.

3. RESULTS

3.1 Description of Experiments

Fifteen Monte Carlo experiments, each of 1800 simulated data sets, were run. The fifteen experiments were composed

of one experiment for each of $n = 5$, 10, and 20 subjects per dose group, using each of five models for the tolerance distribution. In each experiment there were $k = 10$ dose groups at dose levels $x_i = i$, $i = 1,\ldots,10$.

The five tolerance distribution models were all logistic or symmetrically contaminated logistic with $\mu = 5.5$. Denote by $L(\tau)$ the logistic distribution function having mean 5.5 and standard deviation τ. Then $P(\cdot)$ for Model I is the logistic tolerance distribution $L(\tau)$ and the $P(\cdot)$ for Models II, III, and IV are the contaminated tolerance distributions $0.9L(\sigma_c) + 0.1L(c\sigma_c)$, $c = 3$, 5, and 10. The parameters τ, σ_3, σ_5, and σ_{10} were chosen so that $P(\cdot)$ for models I-IV satisfied $P(3) = 0.01$ and $P(8) = 0.99$. Let $D(d)$ be the distribution function of a random variable degenerate at d. Then $P(\cdot)$ for Model V is the degenerate contaminated tolerance distribution, $0.1D(1.5) + 0.8L(\tau) + 0.1D(9.5)$. The numerical values of $P(x_i)$, $i = 1,\ldots,10$, for each model are given in Hamilton (1979).

The Monte Carlo experiments are intended to simulate data from routine lethal toxicity tests on fish and aquatic organisms. Such tests typically involve 4 to 10 dose levels, which are widely spaced because of the experimenter's lack-of-knowledge about the potency of the pollutent being tested. Ten dose levels were used in the simulation design to give the force of contamination a chance to affect the estimates. Because the doses are widely spaced, it is not unusual for partial kills (i.e., $s_i \neq 0$ or 1) to occur at only one, two, or three dose levels (Committee on Methods for Toxicity Tests with Aquatic Organisms (1975); Stephan (1977)).

Specific details concerning the Monte Carlo simulation can be found in Hamilton (1979)).

TABLE I

Relative Bias and Standard Deviation of the Estimated Standard Error and the Empirical Level of Nominal 95 Percent Confidence Intervals for the ED50; The Entries are Based on Monte Carlo Simulation

Estimator	n = 5				n = 10				n = 20			
	Relative Bias (%)[a]	$\sigma(\widehat{SE})$	Confidence Level (%)	No. of Data Sets	Relative Bias (%)[a]	$\sigma(\widehat{SE})$	Confidence Level (%)	No. of Data Sets	Relative Bias (%)[a]	$\sigma(\widehat{SE})$	Confidence Level (%)	No. of Data Sets
Model I: Logistic												
MLE	.9	.45	94.0	1517	- 4.1	.26	94.8	1777	- 1.6	.14	95.1	1799
SK	-10.0	.57	91.9	1761	- 7.9	.30	93.4	1788	- 2.8	.15	94.7	1784
5RSK	- 2.8	.63	94.6	1762	- 3.3	.35	93.7	1788	- 1.3	.18	94.1	1800
10RSK	- .6	.76	93.8	1764	- 3.8	.46	93.5	1799	- .9	.22	93.9	1800
20RSK	- 3.5	1.32	96.7	1764	1.1	.74	96.4	1799	5.1	.40	96.7	1800
Model II: $0.9\,L(\sigma_3) + 0.1\,L(3\sigma_3)$												
MLE	32.3	.47	97.0	1228	- .6	.28	95.7	1704	1.3	.16	95.2	1799
SK	7.7	.58	94.1	1667	- 9.5	.33	92.3	1756	- 2.3	.17	94.3	1695
5RSK	3.4	.63	96.1	1687	- 2.9	.38	93.9	1795	- .9	.21	94.6	1800
10RSK	7.0	.75	95.0	1694	- 3.4	.52	94.9	1796	.6	.27	95.9	1800
20RSK	4.6	1.29	98.6	1694	3.3	.85	98.5	1796	12.1	.53	98.0	1800
Model III: $0.9\,L(\sigma_9) + 0.1\,L(9\sigma_9)$												
MLE	21.1	.48	98.5	866	10.7	.28	97.0	1469	10.8	.16	96.4	1754
SK	2.3	.53	96.6	1451	- 9.6	.37	93.7	1709	- 2.6	.20	94.5	1662
5RSK	14.9	.61	98.4	1479	1.9	.45	97.0	1754	.4	.20	97.3	1797
10RSK	23.5	.76	98.4	1487	- .9	.63	98.7	1756	6.6	.36	99.1	1799
20RSK	14.2	1.15	99.7	1487	5.2	.80	99.9	1756	9.2	.47	99.9	1799
Model IV: $0.9\,L(\sigma_{10}) + 0.1\,L(10\sigma_{10})$												
MLE	32.0	.46	99.7	636	30.4	.26	99.5	1092	35.5	.15	99.4	1560
SK	16.7	.43	99.0	1037	- .9	.33	97.7	1469	- 1.5	.22	97.3	1634
5RSK	52.0	.57	99.7	1065	22.1	.49	99.9	1514	5.4	.32	99.7	1752
10RSK	99.7	.80	99.7	1073	12.5	.52	100.0	1516	12.0	.25	100.0	1752
20RSK	42.0	.75	100.0	1073	18.3	.38	100.0	1516	12.0	.17	100.0	1752
Model V: Degenerate Contaminated Logistic												
MLE	- .4	.81	95.2	1799	- 1.6	.39	94.6	1800	1.0	.19	95.0	1800
SK	- 8.6	.46	93.5	1793	- 6.5	.31	92.5	1791	10.3	.16	95.0	1792
5RSK	- 1.9	.76	94.2	1793	- 3.7	.40	94.4	1795	- 1.2	.21	95.4	1800
10RSK	.3	1.06	94.5	1794	- 4.4	.50	94.1	1800	.5	.31	95.8	1800
20RSK	- 5.1	1.69	95.8	1798	- 2.0	.87	93.4	1800	2.1	.46	95.3	1800

[a] Relative Bias = $[\hat{s}(\widehat{SE}) - \sigma(\hat{\xi})]/\sigma(\hat{\xi})$.

3.2 Comparisons of Standard Error Estimators

The estimate of μ and the estimated standard error of the estimate were calculated by each method for each of the 27,000 (= 15 x 1800) data sets. Let $\bar{E}(\widehat{SE})$ be the empirical mean of the standard error estimate; let $\sigma(\hat{\mu})$ be the empirical standard error of $\hat{\mu}$. The relative bias of the standard error, $[\bar{E}(\widehat{SE}) - \sigma(\hat{\mu})]/\sigma(\hat{\mu})$, and the empirical standard error of $\widehat{SE}$, $\sigma(\widehat{SE})$, were calculated for each method for each experiment. The results are shown in the table.

Some data sets were generated for which one or more of the estimators of μ could not be calculated. (Hamilton (1979) discusses incalculability for each of the estimators.) Each entry in the table is based on all data sets in the experiment for which the associated $\hat{\mu}$ was calculable and $\widehat{SE}(\hat{\mu})$ was greater than zero. The number of such data sets is shown in the table.

Except for the heavy-tailed tolerance distribution models (Models III and IV) none of the standard error estimators exhibit great bias. All negative relative bias is less than 10 percent; all important bias is in the conservative, positive direction.

Typically, MLE has the least variable standard error estimate. The variability of the estimated standard error of the trimmed SK estimate usually increases as the trimming proportion A increases, the main exception being that $\sigma(\widehat{SE})$ can be smaller for 20%SK if the tolerance distribution is heavy-tailed (Models III and IV).

3.3 Comparisons of Confidence Interval Estimates

The empirical confidence levels of the nominal 95 percent confidence interval procedures are shown in the table. The

empirical confidence level for MLE is never below 94 percent, is very close to the nominal 95 percent for light-tailed tolerance distributions (Models I, II, V), and is between 96.4 percent and 99.7 percent for the heavy-tailed tolerance distributions (Models III and IV).

The trimmed SK procedures become increasingly conservative as the trimming proportion A increases. The confidence intervals associated with SK yield empirical levels as small as 91.9 percent. But the 10% and 20%SK confidence intervals generally have levels larger than the nominal 95 percent and, for the heavy-tailed tolerance distribution model (Model IV), the levels reach 100.0 percent.

One could raise the level of the SK confidence intervals by replacing n_i by $(n_i - 1)$ in (2.4). This change forces the empirical confidence level for SK to be greater than 93.7 percent in all of the fifteen experiments. Such a change, however, widens the already conservative 10%SK and 20%SK confidence intervals.

The probability that a confidence interval excludes the value θ is called the accuracy at θ of the interval estimator. To determine the accuracy of each confidence interval procedure, the proportions of times the calculated intervals excluded the (incorrect) ED50 value θ were found for $\theta = 5.6\ (0.2)\ 6.6$. The most accurate estimator will have the largest values for these proportions. The calculated proportions for MLE, SK, 5%SK, and 10%SK are plotted in Figures 1-6. These plots convey the same information as do power plots in hypothesis testing methodology.

For these plots, the same data sets were used to calculate the proportions for all estimation procedures. Thus a data set was used in the calculations only if all of the interval estima-

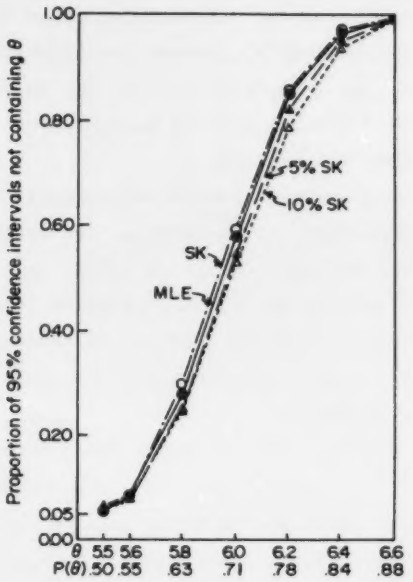

Fig. 1 Model I Logistic;
n = 10; 1766 Data Sets

tors under study were calculable and all $\hat{SE}(\hat{\mu})$'s were greater than zero for that data set. The number of such data sets is given in the label for each figure.

When measuring the accuracy of an ED50 confidence interval, it is useful to determine which percentiles of the underlying tolerance distribution are excluded from the interval. Notice that θ is the $100 \cdot P(\theta)^{th}$ percentile of the tolerance distribution and that the abscissa of each figure is labeled with $P(\theta)$ as well as θ. Thus in Figure 1, $\theta = 5.8$ is the 63^{rd} percentile; 5% and 10%SK excluded this 63^{rd} percentile 25 percent of the time and MLE excluded the 63^{rd} percentile 28 percent of the time.

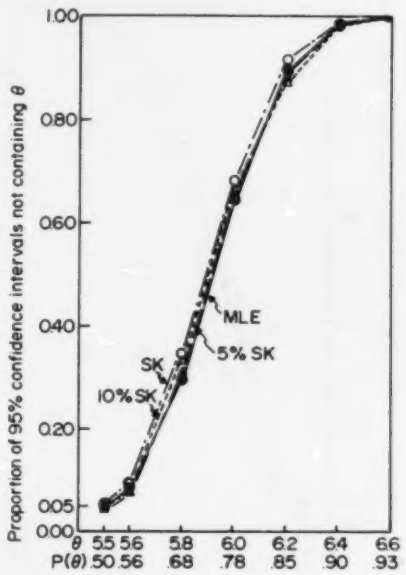

Fig. 2 Model II $.9L(\sigma_3) + .1L(3\sigma_3)$; n = 10; 1664 Data Sets

The accuracy of all interval estimators improves as n increases (e.g., compare Figures 1, 5, 6). The accuracy at n = 5 was so poor for all procedures that I judge them to be of limited value.

Overall, SK and MLE intervals have slightly higher accuracy than trimmed SK intervals for light-tailed tolerance distributions (cf. Figures 1, 2). The 10%SK intervals are much more accurate for the heavy-tailed tolerance distributions (cf. Figures 3 and 4). The logistic MLE is far too conservative when the tolerance distribution is heavy-tailed. The 20%SK intervals are less accurate than the 10%SK intervals

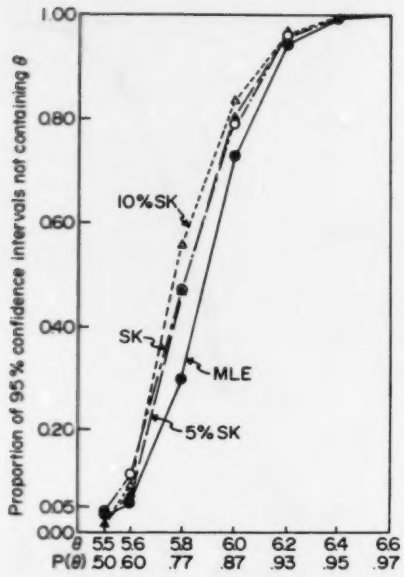

Fig. 3 Model III $.9L(\sigma_5) + .1L(5\sigma_5)$; $n = 10$; 1402 Data Sets

for light-tailed tolerance distributions and very slightly more accurate than 10%SK for the heavy-tailed distributions. All of these statements concerning accuracy hold even if n_i is replaced by $(n_i - 1)$ in (2.4).

4. DISCUSSION

Consider a binary response bioassay design where $P(x_1) \doteq 0$, $P(x_k) \doteq 1$, and the doses are evenly spaced. Among all such designs, at one extreme is the design where there are many subjects at each dose and many doses between $P^{-1}(.15)$ and $P^{-1}(.85)$. For this design it is well-known that the maximum

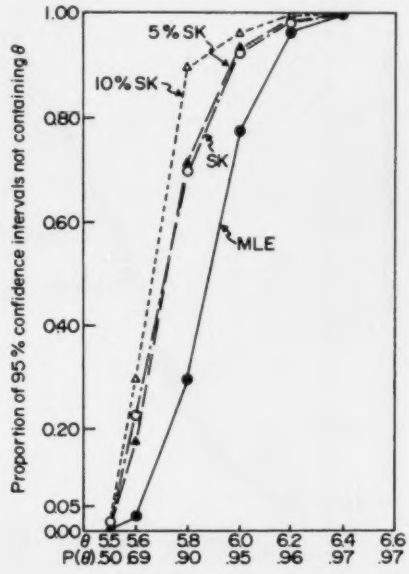

Fig. 4 Model IV $.9L(\sigma_{10}) + .1L(10\sigma_{10})$; n = 10; 1045 Data Sets

likelihood, the Spearman-Kärber, and the trimmed Spearman-Kärber estimators of the ED(50) are approximately normally distributed (Miller and Halpern (1979)). Confidence interval estimates can reliably be based on the asymptotic normal distributions.

At the other extreme are the small, coarse designs used in this simulation study; viz., where there are few subjects at each dose and only a couple of doses between $P^{-1}(.15)$ and $P^{-1}(.85)$. If the confidence intervals based on (2.5) have the correct level and acceptable accuracy for these small designs, then one can reasonably expect the interval estimation procedures to be valid for larger designs. Other reasons for using small, coarse designs in the simulation study are that such designs

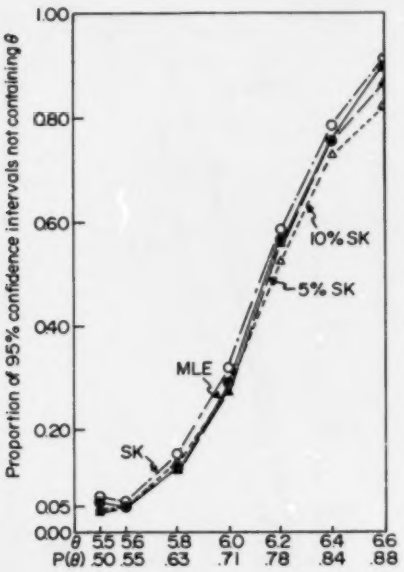

Fig. 5 Model I Logistic; n = 5; 1514 Data Sets

are used in practice (e.g., for routine lethal toxicity studies using aquatic organisms) and that the trimmed Spearman-Kärber estimator appears to be especially useful for analyzing many routine, small bioassays. It would be of interest in future statistical studies, however, to simulate moderate-sized designs with 10-50 subjects at each dose, with five or more doses between $P^{-1}(.15)$ and $P^{-1}(.85)$, and with asymmetric placement of the doses around the ED50.

Because of omission of data sets due to incalculability, the results presented in the preceeding section are based on only a portion of the sample space of $(r_1,\ldots,r_k)$. It is necessary to truncate the sample space for a couple of reasons.

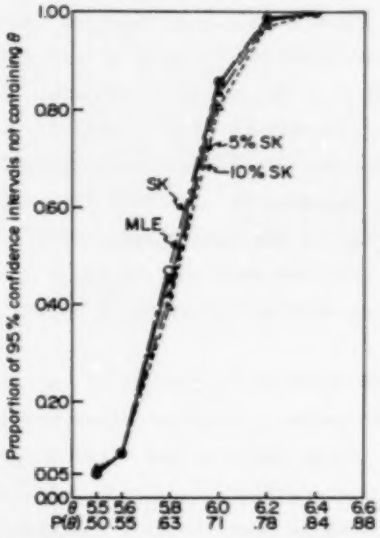

Fig. 6 Model I Logistic;
n = 20; 1783 Data Sets

First, data sufficiently pathological for the experimenter to toss out the whole experiment can occur. For example, even for the widely spaced doses used in this study, there is positive probability that all r_i's = 0.

Second, a specific estimator may fail on a set of potentially informative data. Consider MLE; the maximum likelihood normal equations are functions of two parameters, μ and β^{-1} (which is proportional to the standard deviation of the tolerance distribution). If $r_1 = \ldots = r_{j-1} = 0$, $r_j \neq 0$ or n_j, and $r_i = n_i$ for $i = j+1,\ldots,k$, then the unique solution to the normal equations is $\hat{\mu} = x_j$ and $\hat{\beta}^{-1} = 0$. But if the variance of the tolerance distribution is estimated to be

zero, then the estimated standard error of $\hat{\mu}$ must be zero and the confidence interval must collapse to the single point $\hat{\mu}$. When confronting such data, (not an unusual situation in lethal toxicity bioassays in aquatic toxicology), the experimenter will not be able to use the maximum likelihood method. It seems reasonable to calculate the properties of MLE only over that portion of the sample space where MLE is useful. For this reason, a data set such that $r_i(n_i - r_i) \neq 0$ at 0 or 1 dose level was omitted for purposes of investigating the properties of MLE.

This attempt to describe properties of the estimators in a manner relevant to real experimentation has the disadvantage that the truncated sample space is not clearly defined. One partial solution to this difficulty is to create an *ad hoc* adjustment to an estimator when it isn't calculable. Almost all data sets could then be used to determine the properties of the estimator. For example, one could modify MLE such that the SK point, standard error, and confidence interval estimates are calculated whenever $r_i(n_i - r_i) \neq 0$ at 0 or 1 dose level. This approach has not been explored. In my opinion, such modifications are unlikely to alter the conclusions about the relative value of the trimmed SK confidence interval method.

5. CONCLUSION

The conclusions which follow are based on this Monte Carlo study and are appropriate only for experiments where the dose levels enclose most of the tolerance distribution.

If the tolerance distribution is logistic, the estimated standard error of the trimmed SK estimate is reasonably unbiased and the associated 95 percent confidence interval

has a nearly correct level with accuracy almost as high as with the logistic MLE procedure. For heavy-tailed, contaminated logistic tolerance distributions, the trimmed SK standard errors are positively biased, but the interval estimators are much more accurate than for the logistic MLE procedure.

For experiments with n = 10 or 20 subjects per dose, the author believes that the confidence intervals for SK, 5%SK, 10%SK, and 20%SK are accurate enough to be of practical value. The trimmed SK confidence intervals are especially useful when the underlying tolerance distribution is heavy-tailed.

ACKNOWLEDGEMENTS

Part of this research was performed during 1978 while the author was on assignment to the National Institute of Environmental Health Sciences at Research Triangle Park, North Carolina. This work was funded in part by the U.S. Environmental Protection Agency, Environmental Research Laboratory - Duluth, Research Grant No. R803950. The author is grateful to Dr. M.W.J. Layard for comments and criticism that led to the detection of a major programming error in the first attempt at this Monte Carlo study. The manuscript was typed by Debbie Morris.

BIBLIOGRAPHY

Ashton, E. D. (1972). *The Logit Transformation.* New York: Hafner Publishing Co.

Bennett, B. M. (1952). Estimation of LD50 by Moving Averages. *J. Hygiene 50*, 157-164.

Committee on Methods for Toxicity Tests with Aquatic Organisms. (1975). *Methods for Acute Toxicity Tests with Fish, Macroinvertebrates, and Amphibians.* Corvallis, OR: U.S. Environmental Protection Agency Ecological Research Series, EPA-660/3-75-009.

Cox, D. R. (1970). *The Analysis of Binary Data.* London: Methuen and Co.

Hamilton, M. A., Russo, R. C., and Thurston, R. V. (1977). Trimmed Spearman-Karber Method for Estimating Median Lethal Concentrations in Toxicity Bioassays. *Environmental Science and Technology 11,* 714-719. Correction (1978). *12,* 417.

Hamilton, M. A. (1979). Robust Estimates of the ED50. *J. Amer. Statist. Assoc. 74,* 344-354.

Miller, R. G. Jr. (1973). Nonparametric Estimators of the Mean Tolerance in Bioassay. *Biometrika 60,* 535-542.

Miller, R. G. Jr. and Halpern, J. W. (1979). Robust Estimators for Quantal Bioassay. Tech. Report 42, Div. of Biostatistics, Standford Univ., Stanford, CA.

Stephan, C. E. (1977). Methods for Calculating an LC50, *Aquatic Toxicology and Hazard Evaluation. ASTM STP 634* (F. L. Mayer and J. L. Hamelink, Eds.). Philadelphia: American Society for Testing and Materials, 65-84.

Received June, 1979; Revised August, 1979.

Refereed by William H. Rogers, The Rand Corporation, Santa Monica, CA.